

FRIDAY
MAY 2
2025

2025 BIOSTATISTICS PRACTICUM/APEX SYMPOSIUM



COLUMBIA

MAILMAN SCHOOL
OF PUBLIC HEALTH
BIOSTATISTICS



**Join us in celebrating the
Master's class of 2025!**

Please scan the QR code to see
the full student titles and abstracts

SCHEDULE

TIME	EVENT	LOCATION
12:00 - 12:45	Welcome & Keynote Speaker	Hammer 401
1:00 - 2:00	Session 1	Rooms: HSC 303, 305 HSC LL 103, 106, 107, 108A, 109A, 202, 204, 205, 207, 210
2:00 - 3:30	Break	
2:30 - 3:30	Session 2	
3:30 - 3:45	Break	
3:45 - 4:45	Session 3	

KEYNOTE SPEAKER



Hartaig Singh (Class of 2018)

"Paths & Perspectives"

Hartaig is currently a Staff Data Scientist at Zocdoc, where he leverages statistics and machine learning to drive strategic marketing insights. Since earning his MS in Biostatistics Theory & Methods from Columbia University's Mailman School of Public Health in 2018, he has built a diverse career spanning

consulting, healthcare, pharmaceuticals, and tech startups. His work has consistently focused on data science with a specialization in marketing analytics and enjoys combining rigorous statistical methods with real-world business impact.

ROOM TOPICS

SESSION 1

Environmental Health Research	MPH	HSC 303
Data Analysis	MPH	HSC 305
Clinical Trials	MPH	HSC LL 103
Cancer Research	MS	HSC LL 106
Deep Learning	MS	HSC LL 107
Longitudinal Data Analysis	MS	HSC LL 108A
Missing Data	MS	HSC LL 109A
Topics in Mental Health	MS	HSC LL 202
Drug Development	MS	HSC LL 204
Data Visualizations	MS	HSC LL 205
Data Analysis	MS	HSC LL 207
Epidemiology	MS	HSC LL 210

SESSION 2

Data Analysis	MPH	HSC 303
Topics in Mental Health	MPH	HSC 305
Epidemiology	MPH	HSC LL 103
Machine Learning	MS	HSC LL 106
Statistical Genetics	MS	HSC LL 107
Bayesian Statistics	MS	HSC LL 108A
Causal Inference	MS	HSC LL 109A
Machine Learning	MS	HSC LL 202
Cancer Research	MS	HSC LL 204
Statistical Genetics	MS	HSC LL 205
Causal Inference	MS	HSC LL 207
Observational Study	MS	HSC LL 210

SESSION 3

COVID-19 Research	MPH	HSC 303
Data Visualizations	MPH	HSC 305
Clinical Trials	MS	HSC LL 103
Cancer Research	MS	HSC LL 106
Functional Data Analysis	MS	HSC LL 107
Machine Learning	MS	HSC LL 108A
Environmental Health Research	MS	HSC LL 109A
Longitudinal Health Research	MS	HSC LL 202
Survey Statistics	MS	HSC LL 204
Epidemiology	MS	HSC LL 205
Observational Study	MS	HSC LL 207
Multi-Interest Session	MS	HSC LL 210

SESSION 1

HSC 303 <i>Environmental Health Research</i> MPH Host: Safiya Sirota	Qianxuan Huang Predicting General Health Outcomes Using Environmental and Socioeconomic Factors: A Longitudinal Analysis of BRFSS Data Xiwen Wu Water Quality in Public Health Shizhe Zhang Greenspace and Health Equity in the South Bronx: Assessing Environmental Disparities and Public Health Impacts Aakansha Bagepally Associations of Hair Product Use, Women's Health, and Reproductive Outcomes
HSC 305 <i>Data Analysis</i> MPH Host: Yuying Lu	Yu Huang Choice Of Metric And The Effect Of Scan Length For Reliability In Resting-State fMRI Tianqi Li Analysis And Interpretation Of Population Differences in Structural Variation Of The Human Genome Coco Ni Enhancing Data-Driven Decision-Making for Long-Acting Injectable HIV Treatment: Development of an Interactive R Shiny Dashboard Xin Hui Population Differences In Gene Expression And Resulting Differences In Disease Incidence
HSC LL 103 <i>Clinical Trials</i> MPH Host: Gustavo Garcia-Franceschini	Enju Zhang Building SDTM & ADaM for Regulatory Compliance Yuanyu Lu Effects Of Transcranial Alternating Current Stimulation On Measures Of Cognition And Symptom Scores In Chinese Patients With Schizophrenia Siqi Wang Site Selection Tool For Analysts
HSC LL 106 <i>Cancer Research</i> MS Host: Ryan Wei	Angela Dauti Exploring Real-World Data for Clinical and Pharmacological Use Cases: A Tumor-Specific Analysis in Oncology Shuyan Qiu Real-world Cancer Treatment Patterns And Their Effects In Older Adults With Lung Cancer: A SEER-MHOS Analysis Haochen Shi Examining Allostatic Load Disparities Across Age and Demographic Factors in Breast and Digestive Cancer Patients Siqing Wang Disparities In Treatment Delays Among Patients With Pancreatic Adenocarcinoma: A SEER-Medicare Study
HSC LL 107 <i>Deep Learning</i> MS Host: Shubo Zhang	Yimeng Cai Automated Paper Screening for Clinical Reviews Using Large Language Models: Data Analysis Study Siyan Wen The Impact of Image Quality on Prediction Performance: A Study Using Alzheimer's Disease Neuroimaging Initiative Data Ruiqi Xue Deep Learning For Identifying Key Brain Features Associated With Apathy In Alzheimer's Disease
HSC LL 108A <i>Longitudinal Data Analysis</i> MS Host: Iris Wang	Luxuan Zhang Evaluating The Effectiveness Of Asthma Self-Management For Adolescents (ASMA) In Rural Areas: A GEE-Based Analysis Yuqi Cheng Longitudinal Associations of Uric Acid with Metabolic Parameters - A Mixed-Effects Analysis in Chinese Patients with Metabolic Disorders Ruoying Deng Analysis of Medication Effects on Cognitive Outcomes in Adni Data Yuzhe Hu Biomarker Trajectories in Alzheimer's Disease: A Semiparametric Joint Modeling Approach and Its Application

SESSION 1

HSC LL 109A <i>Missing Data</i> MS Host: Bin Yang	Xiaoting Tang Quantification of Biological Aging for Prognosis in ICU Patients Eunice Wang Sepsis Treatment Bundle Complicance Derek Lamb Leveraging trans-pQTLs to Improve Protein Prediction in Biobank-scale Proteome-wide Association Studies Chenshuo Pan Sepsis Treatment Analysis
HSC LL 202 <i>Topics in Mental Health</i> MS Host: Yanran Li	Shuchen Dong Resilience to suicide ideation in high risk adolescents: the role of physical activity Yilei Yang Factors Influencing Workforce Retention Among Foreign Trained Nursing Workforce: A National Sample Survey Analysis Wenyu Zhang Machine Learning Identification of Thalamic Biomarkers for Early Cognitive Impairment in At-Risk MS Mengxiao Luan Harmonizing Two Likert Variables with Identical Questions but Different Response Scales
HSC LL 204 <i>Drug Development</i> MS Host: Yiming Li	Chen Liang Efficacy of Batoclimab in gMG: Correlation of IgG and AChR-Abs with Clinical Outcomes Shaolei Ma R You Ready? Developing R-Based Tools for TFL Review in Clinical Trials Yuki Low Statistical Designs for Rare Disease Trial Proposals to improve Clinical Trial Readiness for Regulatory Approval: An Update Zihan Wu A Comprehensive Meta-Analysis of Phase 3 Randomized Controlled Trials for Alzheimer's Disease Treatments
HSC LL 205 <i>Data Visualizations</i> MS Host: Kate Colvin	Sitian Zhou Understanding TB Symptoms And Household Transmission In Uganda Huanyu Chen Spatial Heterogeneity In COVID-19 Response: Leveraging Regression Models To Explore Socioeconomic And Demographic Disparities In New York City Jiatong Li Cognitive Trajectories Across Education Levels and Gender: A Longitudinal Analysis of the Health and Retirement Study Zixuan Qiu Construction Of Clinical Trial data Integration Platform and Review Of CSM Methodology Framework
HSC LL 207 <i>Data Analysis</i> MS Host: Zexi Cai	Ziqi Liao BLOOM: Robust Maximin Optimization for Data Fusion with Blockwise Missing Covariates Kairui Wang Metabolomic and Proteomic Analyses of Plasma in Gulf War Illness before and after Exercise Miao Fu Understanding the Mismatch Between Self-Reported and Actual HIV Status: Predictors of Unawareness Among HIV-Positive Individuals Allison Xia An Interactive Model Performance Assessment for Advanced NSCLC Diagnosis
HSC LL 210 <i>Epidemiology</i> MS Host: Ting-Hsuan Chang	Justin Vargas Examining The Relationship Between Home And School Social And Environmental Exposures And Childhood Lung Function Outcomes Xue Zhang Epidemiologic Analysis of COVID-19 and Gastrointestinal Outbreaks in Pima County, Arizona (2023) Longyi Zhao The Association Between Weight Changes Over Time and Pancreatic Cancer Risk in a Pooled Analysis of 8 Prospective Cohort Study in the Pooling Project of Prospective Studies of Diet and Cancer (DCPP)

SESSION 2

HSC 303 <i>Data Analysis</i> MPH Host: Safiya Sirota	Li Jiang Leveraging Data Analytics for Biotech Operations and Market Strategy Huizhong Peng Lifestyle Risk Factors in Middle-aged and Older Adults: Insights from the UK Biobank Study Yiru Xiang Analyzing Gut Microbiome Dynamics in Intestinal Transplant Recipients: Investigating Its Role in Graft Rejection Chengyuan Zhang Data Processing And Data Analysis Regarding Medical Funds: Find Indexes That Correlate With Medical Funds' Performance in China in the Post-Covid Era
HSC 305 Topics in <i>Mental Health</i> MPH Host Yuying Lu:	Jiayi Li Associations Between Maternal Allostatic Load and Placental DNA Methylation: Targeted Gene and Epigenome-Wide Analyses Ruiyang Wu Prenatal Social Determinants of Health and Mental Health and Newborn Birth Outcomes and Feeding Jiayi Zhang Anxiety Disorder Analysis based on the Generalized Anxiety Disorder-7 Survey Xuezheng Wang Demographic Disparities in Youth Psychiatric Inpatient Discharges: Evidence from New York State SPARCS Data (2019–2022)
HSC LL 103 <i>Epidemiology</i> MPH Host: Gustavo Garcia-Franceschini	Hongyi Chen Identification of Preeclampsia Subtypes Via Machine Learning and Associations With Risk of Perinatal and Postpartum Outcomes Katherine Gong Association Between Pre-Pregnancy Magnesium Intake and Preterm Birth: Evidence from the NuMoM2b Prospective Cohort Study Sining Leng Evaluating the Impact of Elder Justice Services Through Data Analysis and Public Health Strategies
HSC LL 106 <i>Machine Learning</i> MS Host: Ryan Wei	Tianhui Huang Novel Pipelines to Extract Differences in Proteome Dynamics Based on Health Status Zhi Heng Shi Anti-Sepsis for Infants and Moms and Its Application Yuntian Xu Machine Learning Module in EasyR for Clinical Data Analysis Qiran Chen Supervised Machine Learning Analysis for Prostate Cancer MRI Imaging
HSC LL 107 <i>Statistical Genetics</i> MS Host: Shubo Zhang	Shihang Zeng Exploring Quantile-Specific Genetic Associations with Alzheimer's Disease Using Quantile TWAS Aiying Huang MS-COEX: Multi-sample Cell-type-specific Coexpression Detection in Single-cell RNA-seq Data Xinyi Shang Identifying Cell-Type-Specific Spatially Variable Genes with ctSVG Joshua Carpenter Multi-Omics Insights into Maternal Obesity and Fetal Metabolism
HSC LL 108A <i>Bayesian Statistics</i> MS Host: Iris Wang	Iris Yang Bayesian Hierarchical Modeling of U.S. Death Rates During 1973-2022 Alexander McCreight SuSiE.ash: A Sum of Single Effects Regression and Adaptive Shrinkage Model as a Unified Framework for Fine-mapping and TWAS Kindle Zhang Comparing Longitudinal Data Analysis Methods with Simulation Data

SESSION 2

HSC LL 109A <i>Causal Inference</i> MS Host: Bin Yang	Sixuan Chen Improving Indirect Treatment Comparisons Using Matching-Adjusted Indirect Comparisons (MAIC) Ziqiu Liu The Role Of Mitochondrial Energy Expenditure In Brain And Cognitive Aging Yumeng Qi Longitudinal Relationships Between REM Sleep Behavior Disorder, Dopaminergic Dysfunction, and Cognitive Decline in Parkinson's Disease Yixiao Sun Enhancing Comparative Effectiveness Research with Matching-Adjusted Indirect Comparison: A Robust Approach for Treatment Comparisons
HSC LL 202 <i>Machine Learning</i> MS Host: Yanran Li	Ru Jin Lim HRSA Health Center Workforce Wellbeing Survey Analysis Jiarui Yu Actuarial Mortality Modeling: Enhancing Accuracy with Optimized Rates Wenxin Tian Differential Enrichment Analysis of Antibody Responses Among Monkeypox Survivors Using Phage Display Yuxuan Du Trajectory Analysis For Multiple Sclerosis Patients
HSC LL 204 <i>Cancer Research</i> MS Host: Yiming Li	Lu Qiu Global, Regional, and National Burden of Male Breast Cancer, 1990–2021, with Projections to 2050: A Systematic Analysis of the Global Burden of Disease Study 2021 Yujia Zhang Survival Model: Ovarian Cancer Patient Survival Measured with Treatment Data Yang Zhao Assessing The Efficacy Of Dendritic Cell-Based Vaccine Therapy Mingzhi Chen Differential Expression Analysis of Glioblastoma Cells
HSC LL 205 <i>Statistical Genetics</i> MS Host: Kate Colvin	Lehan Zou LamianOmni: Extending Multi-Sample Pseudotime Analysis to DNA Methylation Data Jiying Wang Lineage Tracing of Thymic Epithelial Cells Using Single-Cell Multi-omics Yunshen Bai Regulatory Networks and Gene Expression: A Comprehensive Simulation Study Jingyi Xu Evaluating Control Group Composition In Case-Control Studies: The Role Of Family History In Genetic Risk Assessment
HSC LL 207 <i>Causal Inference</i> MS Host: Zexi Cai	Yifei Liu Exploring the Causal Pathway from Blood Pressure and Stroke to Alzheimer's Disease: A Mendelian Randomization Study Ixtaccihuatl Obregon Adaptation Of An Omnibus Test To Detect Subgroup Treatment Effects For Observational Studies Hongzhu Ren Exploratory Analysis of Causal Relationships Among 19 Inflammaging Biomarkers Zhiyi Zhu Bias Correction in PET Kinetic Modeling: Refining the Simex Algorithm for Noisy Reference Inputs
HSC LL 210 <i>Observational Study</i> MS Host: Ting-Hsuan Chang	Dylan Koproski Analyzing Pre-Pandemic Age-Based Visitation Trends Using Advan Cell Phone Tracking Data Shaoyu Chen Identifying Predictors of Survival and the Impact of Pretransplant Comorbidities on Pediatric Allogeneic Hematopoietic Cell Transplantation Outcomes Jake Coldiron Building and Validating a Federated Learning Algorithm with Virtual Pooling for Rare Genetic Diseases Ruoxi Li Enhancing Predictive Accuracy in Small Subgroups Through Fairness-Constrained and Transfer Learning Methods

SESSION 3

HSC 303 COVID-19 Research MPH Host: Safiya Sirota	Chenyu Jin Network and Regression Analysis of U.S. College Student Mental Health and Coping Strategies During COVID-19 Yanlin Cui The Impact of Social Vulnerability and Childhood Opportunity on Child Traffic Mortality in New York State: A Pre- and Post-COVID Comparison Zhangfan Xia Infant Feeding in the COVID-19 Mother-Baby Outcomes Study
HSC 305 Data Visualizations MPH Host: Yuying Lu	Xinghao Qiao Potential of DTP Pharmacies Huachen Shan China Pharmaceutical Market Overview and Trends Tiancheng Shi Visualizing Hurricane Trends With An Interactive R Shiny Map Sanne Glastra Climate Resilience Dashboard for Mathematica Policy Research
HSC LL 103 Clinical Trials MS Host: Gustavo Garcia-Franceschini	Yizhuo Chang The Efficacy of Computerized Cognitive Training in Patients with Coronary Heart Disease and Mild Cognitive Impairment Yangyang Chen Revolutionizing Clinical Decision-Making: A Natural Language Chatbot for Clinical Trials Data Retrieval Zhesu Jiang Enhancing MOUD Treatment Insights: Identifying Underrepresented Subgroups and Estimating Conditional Treatment Effects Using Superlearner Lauren Lazaro Optimizing Clinical Trial Design for Polypill Research
HSC LL 106 Cancer Research MS Host: Ryan Wei	Cynthia Cui Evaluating the Impact of Medicaid Expansion on Access to NGS Testing for Advanced Cancer Patients: A Difference-in-Differences Analysis Mirah Koota Distribution of Early-Stage NSCLC Cases and Surgical Resections in the USA using SEER Database for Merck KN-671 Yueyi Xu Survival Outcomes Of Neoadjuvant, Adjuvant, And Perioperative Immunotherapy In Resectable Non-small Cell Lung Cancer: An Analysis Of The National Database Haotian Tang Integrative Multi-Omics Approach Using Augmented Similarity Network Fusion For Cancer Subtype Identification And Clinical Validation
HSC LL 107 Functional Data Analysis MS Host: Shubo Zhang	Peng Su Ecological Scaling Of Temporal Fluctuations With Bacterial Abundance In Gut Microbiota Depends On Functional Properties Of Individual Microbial Species Li Tian Echocardiographic Changes in Sickle Cell Disease: Longitudinal Trends and Sex-Specific Differences Yuxin Yin Implementation of a Transcriptome-Wide Association Study (TWAS) Pipeline Using Summary Statistics
HSC LL 108A Machine Learning MS Host: Iris Wang	Manye Dong Predicting Investigator Performance for Efficient Clinical Trial Enrollment Haitian Huang Predicting Sleep Disorders Using Biostatistical and Machine Learning Methods Qin Huang Identifying Key Nonverbal Communication Features to Classify Autism in Minimally Verbal Children Zhuodiao Kuang Leveraging Multi-Source Summary-Level Data: Enhanced Risk Predictio in Underrepresented Population Via Transfer Learning

SESSION 3

HSC LL 109A <i>Environmental Health Research</i> MS Host: Bin Yang	Tianyuan Deng Effects of Prenatal Metal Exposure on Birth Outcomes: Evidence from HHEAR Project Data Ruijie He Interactive Dashboard Development for the TEDDY Study: Enhancing Analysis of Environmental Determinants of Diabetes in Young Individuals Nisha Lingam Evaluating the Joint Effects of O3, PM2.5, and NO2 as an Exposure Mixture on Mortality Yuhan Wang The Impact of Wildfire-Specific PM2.5 on Mortality in New York Counties 2006–2016
HSC LL 202 <i>Longitudinal Health Research</i> MS Host: Yanran Li	Jennifer Li Exploring The Impact of Periodontitis on Neurological Function Jixin Li Modeling Parent Fraction In PET Imaging Arthur Starodynov NYC's Salary Transparency Law Bridging The Gender Divide Longyu Zhang Age-Related Changes In Brain Dynamic Functional Connectivity Based On Persistent Homology
HSC LL 204 <i>Survey Statistics</i> MS Host: Yiming Li	Ou Sha The Impact Of Updated Federal Guidelines For Collecting Race And Ethnicity Data Tongxi Yu Multi-dimensional Proteomic Analysis for Age-related Biomarker Discovery Mengyuan Yu Association Between Time-Restricted Eating And Dental Caries Junyi Ren Measuring Tick Risk Along an Urbanization Gradient
HSC LL 205 <i>Epidemiology</i> MS Host: Kate Colvin	Wenwen Li Chikungunya Vaccine Allocation Qianying Wu The Impact of Adverse Childhood Experiences on Cognitive Function in Older Adults Linshen Cai Impact of Low-Molecular-Weight Heparin on Post-Hepatectomy Liver Failure: A Retrospective Cohort Study
HSC LL 207 <i>Observational Study</i> MS Host: Zexi Cai	Yaduo Wang Observational Study To Assess The Overall Treatment Effect Of AD Medication On Essential Cognitive Outcomes Ekaterina Hofrenning Blood-based Biomarkers for Predicting Vascular Contributions to Cognitive Impairment and Dementia Brain Changes Yuki Joyama The Role of Menopausal Status in Exercise-Driven Cognitive Benefits Tianyou He Investigating the Association Between BMI and HbA1c Using Observational Data and Statistical Modeling
HSC LL 210 <i>Multi-Interest Session</i> MS Host: Ting-Hsuan Chang	Zilin Huang Automating Survival Analysis for Clinical Trial Studies Ze Li Investigating Gene-Environment Interactions in Age-Related Hearing Impairment Using a Case-Only Approach

ACKNOWLEDGEMENTS

Thank you to the Symposium Planning Committee!

PhD Student Members:

Ting-Hsuan Chang, Yijin Wang, Yanran Li,
Guadalupe (Lupe) Antonio Lopez

Faculty Members and Staff:

Christine Mauro, Paul McCullough, Lucia Li, Erin
Elliott

Session Chairs:

Safiya Sirota, Yuying Lu, Gustavo Garcia-
Franceschini, Ryan Wei, Shubo Zhang, Iris Wang,
Bin Yang, Yanran Li, Yiming Li, Kate Colvin, Zexi Cai,
Ting-Hsuan Chang

2025 BIOSTATISTICS PRACTICUM/APEX SYMPOSIUM



COLUMBIA

MAILMAN SCHOOL
OF PUBLIC HEALTH

BIOSTATISTICS